An online tool for valuing people's health, including valuing 'dead'

Paul Hansen, Franz Ombler & Trudy Sullivan
paul.hansen@otago.ac.nz, franz@1000minds.com, trudy.sullivan@otago.ac.nz

How is your health today? Are you in any pain? Anxious or depressed? Any problems walking around or looking after yourself? Are you able to perform your usual activities? And, on a scale from 0 to 1, where 0 = dead and 1 = perfect health, how would you rate your *health-related* quality of life (HRQoL)?

As explained in more detail later below, measuring a patient's HRQoL is useful for assessing how unwell they are and also their benefit from being treated. More than a dozen systems for representing HRQoL, known as 'health descriptive systems', are available, of which the EuroQol Group's EQ-5D is one of the most widely used systems in New Zealand and internationally.[1]

EQ-5D

The EQ-5D represents HRQoL on five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Two versions of the EQ-5D exist, differentiated by how many levels they have on each dimension: (1) the EQ-5D-3L, with three levels (Brooks 1996); and its successor introduced in 2009, the EQ-5D-5L, with five levels (Herdman et al. 2011). The EQ-5D-5L is presented in Figure 1. *(How would you represent your HRQoL on the EQ-5D-5L; i.e. which level for each dimension best describes your health today?)*

The EQ-5D-5L, whose five dimensions have five levels of performance each, is capable of representing 3125 (i.e. 5^5) health states – i.e. all combinations of the levels on the five dimensions. Each health state can be denoted by a 5-digit number relating to the relevant level for each dimension listed in Figure 1; e.g. 11111 = no problems on all dimensions, 55555 = extreme problems on all dimensions, etc.

VALUING HEALTH STATES

In addition to being able to *represent* 3125 different health states using the EQ-5D-5L, it's useful to *value* the states too. A 'value set' consists of a value for each state (all 3125 of them), with most values ranging between 1 for full health (no problems on the five dimensions: 11111) and zero for 'dead', with negative values for states worse than dead.

Thus, for example, we would be interested to know what is the value for state 22222 (slight problems on all dimensions) – which for most people is likely to lie somewhere in the range 0-1 (between dead = 0 and 11111 = 1). And so on for the other health states.

Value sets are useful for calculating 'Quality-Adjusted Life Years' (QALYs).[2] QALYs are used in economic evaluations whereby the costs and benefits of various types of spending on health procedures, pharmaceuticals, devices, equipment, etc are compared. For example, NZ's Pharmaceutical Management Agency (PHARMAC) calculates QALYs (and 'cost per QALY' estimates) in order to evaluate the pharmaceuticals it's considering buying on behalf of all citizens.

Another use of value sets is for calculating 'patient-reported outcome measures' (PROMs) for the purpose of evaluating the effectiveness of treatments (often for individual patients).

Figure 1: EQ-5D-5L health descriptive system

DIMENSION
Mobility
I have no problems in walking about
I have slight problems in walking about
I have moderate problems in walking about
I have severe problems in walking about
I am unable to walk about
Self-Care
I have no problems washing or dressing myself
I have slight problems washing or dressing myself
I have moderate problems washing or dressing myself
I have severe problems washing or dressing myself
I am unable to wash or dress myself
Usual Activities (e.g. work, study, housework, family or leisure activities)
I have no problems doing my usual activities
I have slight problems doing my usual activities
I have moderate problems doing my usual activities
I have severe problems doing my usual activities
I am unable to do my usual activities
Pain/Discomfort
I have no pain or discomfort
I have slight pain or discomfort
I have moderate pain or discomfort
I have severe pain or discomfort
I have extreme pain or discomfort
Anxiety/Depression
I am not anxious or depressed
I am slightly anxious or depressed
I am moderately anxious or depressed
I am severely anxious or depressed
I am extremely anxious or depressed

1 EQ-5D stands for EuroQoL, 5 Dimensions. Other health descriptive systems include the HUI (Health Utilities Index), SF-6D (Short Form, 6 Dimensions), 15D (15 Dimensions) and AQoL (Assessment of Quality of Life).

2 As an example of how QALYs are calculated, imagine an elderly woman who needs a hip replacement and is expected to live a further 10 years, but her sore hip reduces her (health-related) quality of life such that each of those years is considered to be worth just 0.6 of a year of 'full health' (valued at 1 QALY). Her *quality-adjusted* life expectancy is therefore 10 × 0.6 = 6 QALYs. After the hip replacement operation, her life expectancy is unchanged, but suppose her quality of life rises from 0.6 to 0.9. She now has 10 × 0.9 = 9 QALYs, and so the surgery produced a gain of 9 – 6 = 3 QALYs.

Value sets are usually created at the population level, intended to represent a population (e.g. New Zealand) 'on average'. Such *social* value sets are usually created using online questionnaires, often supported by specially-trained interviewers, administered to a large sample of the population (e.g. more than 1000 people). Accordingly, producing a social value set is usually an expensive and time-consuming exercise.[3]

NOW IT'S PERSONAL!

As well as a *social* value set for New Zealanders as a whole (on average), it would be ideal to be able to create *personal* value sets for individual patients – i.e. one value set for each patient, based on their personal preferences.

Doing so would enable the fact that people have different preferences with respect to how they feel about the HRQoL dimensions to be more systematically recognised. For example, some people care about pain more than other dimensions, other people care most about being able to walk around, others to being able to care for themselves, etc.

Being able to incorporate a person's HRQoL preferences into decisions about the best treatment for them – from all possible treatments – is consistent with 'personalised' or 'precision' medicine whereby treatments are tailored to the individual based on their risk of disease or predicted treatment response (Mirnezami, Nicholson & Darzi 2012). Enabled by advances in diagnostic approaches, especially genomics, personalised medicine has so far focused on what is *technically possible* – without systematically including information about the patient's *preferences* (until now!).

This article reports on a new online tool for creating *personal* and *social* value sets quickly and relatively cheaply. All the person has to do is spend about 5-10 minutes answering some simple questions and then the tool generates their own EQ-5D-5L value set, as well as contributing to a social value set for the group of participants overall.

The tool can be seen as a breakthrough in the HRQoL value set 'industry'. Its user-friendliness and online delivery could massively reduce the cost of producing value sets for the EQ-5D-5L – or for any other health classification system (see footnote 1).

The tool can be deployed 'in the field' to quickly and cheaply generate the HRQoL preference data required to produce value sets at the population (social) level.

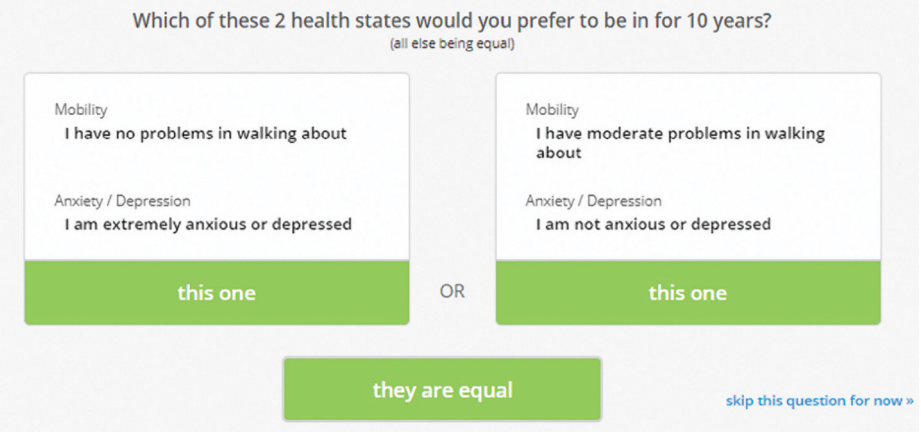
The tool could also be available on computer tablets in doctor waiting rooms or as a mobile app for patients to quickly create their own personal value sets. The easy availability of personal value sets opens up the possibility of individual patient preferences being more systematically incorporated into treatment decisions.

'PAIRWISE RANKING' AND 'BINARY SEARCH'

Powered by 1000minds software (www.1000minds.com), the tool has two main components: (1) a pairwise ranking exercise to determine the 3125 health state values for each participant, and (2) a binary search algorithm to identify any health states they consider to be worse than dead. Both components are described below, but if you would rather experience the tool immediately, please jump to the link to the tool in the second-last section below.

The pairwise ranking exercise is based on the PAPRIKA method (Hansen & Ombler 2008) – an acronym for ***P***otentially ***A***ll ***P***airwise ***R***an***K***ings of all possible Alternatives.[4] In the present context, PAPRIKA involves the participant being repeatedly asked to choose between two hypothetical health states defined on just two dimensions at a time with respect to which state they would prefer to be in for 10 years. Each choice involves a trade-off between the levels for the two dimensions, where implicitly the levels on the other three dimensions are the same for both states (i.e. "all else being equal"). An example of a pairwise-ranking question appears in Figure 2.

Figure 2: Example of a pairwise-ranking question from the 1000minds software



Such simple questions are repeated with different pairs of hypothetical health states – always involving trade-offs between different combinations of attributes, two at a time – until enough information about the person's preferences has been collected to determine their weights on the attributes, thereby generating a value set for that person.

Central to the PAPRIKA method is that it learns from and *adapts* to each person's preferences. Each time a person ranks a pair of health states, all other states that can be pairwise ranked via the logical property of 'transitivity' are identified and eliminated, thereby minimising the number of questions asked.

For example, if a person prefers health state *A* to *B* and *B* to *C*, then – by transitivity! – *A* is also preferred to *C* (and is not asked about). Each time a person answers a pairwise-ranking question, based on all preceding answers, PAPRIKA adapts with respect to choosing the next question (always one whose answer is not implied by earlier answers).

3 The expense involved in creating social value sets is one reason why the value set for the EQ-5D-3L version of the EQ-5D (with just three levels on each dimension instead of five, as for the EQ-5D-5L) created in 1999 (Devlin, Hansen, Kind & Williams 2003) has not been updated since then (until now!).

4 Since 2004, this method and 1000minds software have been used in a wide range of health applications: health technology prioritisation (e.g. Martelli et al. 2016, Sullivan & Hansen 2017), patient prioritisation (Fitzgerald et al. 2011, Hansen, Hendry, Naden, Ombler & Stewart 2012), disease classification and diagnosis (Shiboski et al. 2017) and prioritising diseases for R&D (Tacconelli et al. 2018).

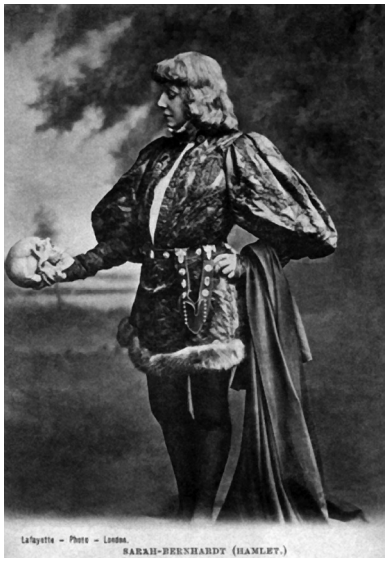
PAPRIKA's adaptivity ensures that the number of questions a person is asked is minimised while ensuring they end up having pairwise ranked all possible health states defined on two dimensions at a time, either explicitly or implicitly (by transitivity).[5] Most people need to answer just 20 questions on average, taking 5-10 minutes in total.

Also, to check the quality of each person's data, two or three repeated questions can be included to assess the person's consistency. Checks can also be made of how long the person took for their answers and for any other evidence that they answered questions unreliably.

TO BE, OR NOT TO BE

Enabled by the existence of a personal value set for each individual, the 1000minds software implements an 'interactive binary search' (or bisection) algorithm for people to identify any health states they consider to be worse than dead. The algorithm developed for the tool is explained in detail here because it is novel compared to more traditional implementations of such algorithms.

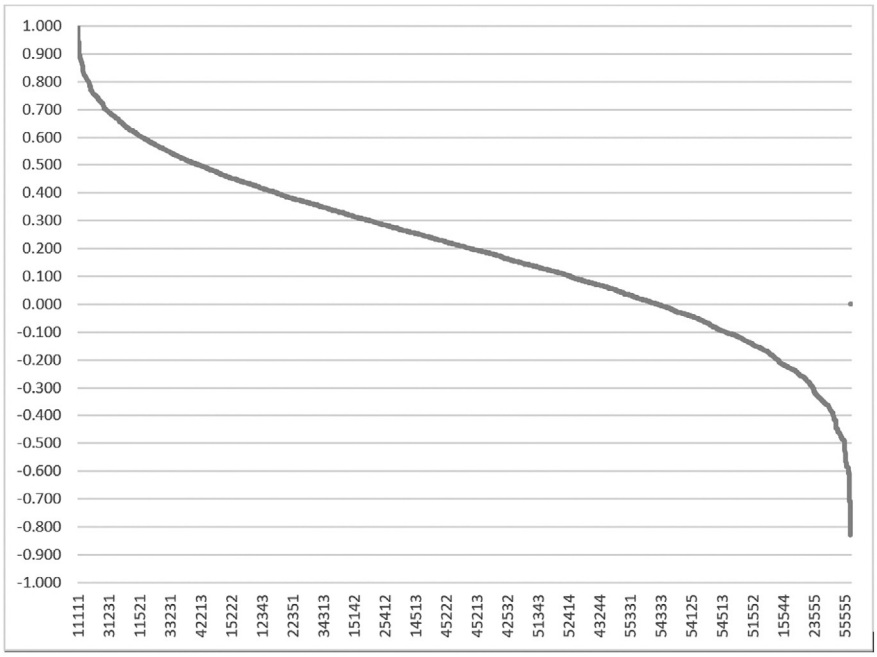
Prince Hamlet: "To die, to sleep, perchance to Dream; aye, there's the rub"



The binary search algorithm begins with the participant being asked if they think that being in the lowest-ranked heath state, 55555 (extreme problems on all dimensions), for 10 years would be better than dead (BTD) or worse than dead (WTD); this question is shown in Figure 3. If the person answers 55555 is BTD, the algorithm stops. If instead they answer 55555 is WTD, the algorithm proceeds to search for, in effect, the 'dividing line' that splits their ranking of the 3125 states into ones BTD and WTD respectively.

Figure 3: Example of a binary search question to identify states worse than dead

Figure 4: The 3125 health state values (means), from highest (11111=1) to lowest (55555=−0.830)



Thus, if the person answers 55555 is WTD, they are asked if another, higher-ranked health state – set by the tool to 33333 (moderate problems on all dimensions) – is BTD or WTD. Depending on their answer, another higher- or lower-ranked state is evaluated: if 33333 is WTD, 22222 (slight problems on all dimensions) is posed next; instead if 33333 is BTD, 44444 (severe problems on all dimensions) is posed next. Having identified the range of health states in which dead lies, the algorithm proceeds to repeatedly bisect (halve) the participant's personal ranking of states.

For example, with reference to the questions above, suppose the person answers 33333 is BTD and then 44444 is WTD; they are then asked if the state in the middle of *their* ranking of 33333 to 44444 – e.g. perhaps 34432 (it depends on *their* ranking) – is BTD or WTD. Suppose 34432 is BTD; they are then asked if the state in the middle of their ranking of 34432 to 44444 – e.g. perhaps 44433 – is BTD or WTD. This process continues: repeatedly halving the range of values until the dividing line is found that splits their ranking of the 3125 states into ones BTD and WTD respectively.

5 To further reduce the number of questions asked, only levels 1, 3 and 5 of each of the EQ-5D-5L dimensions were included in the pairwise ranking exercise. The weights for levels 2 and 4 are interpolated using Bézier interpolation (Farin, Hoschek & Kim 2002) – in essence, fitting a monotonic smoothed curve through the weights for levels 1, 3 and 5. Also, five combinations of levels (health states) deemed to be unrealistic to most people were suppressed: e.g. "no problems doing my usual activities" and either "extreme pain or discomfort" or "extremely anxious or depressed" or "unable to wash or dress myself" or "unable to walk about".

In summary, three results with respect to the location of dead within the 3125 health states are possible for each participant: either dead is worse than 55555, and so dead and 55555 are both valued at 0 (customary for EQ-5D valuations); or 11111 is WTD (uncommon), and so dead = 1; or (most often) dead is spanned by two adjacent states in the person's ranking (one BTD, the other WTD), and so dead's value (before rescaling) is the average of these two states' values.

NEW ZEALAND SURVEY

A sample of the NZ adult population, representative with respect to age, gender, ethnicity and geographic location, was recruited. After extensive checks of the quality of participants' data, a high-quality sub-sample of 2468 people was chosen from which a social value set for NZ was created (as well as 2468 personal value sets). This value set is summarised graphically in Figure 4, where 780 (25%) of the 3125 states are worse than dead.[6]

GO ON, GIVE IT A WHIRL!

To experience the tool yourself – and generate your own EQ-5D-5L value set of 3125 values – open this link: www.1000minds.com/go/eq5d-interpolation-test.[7]

WHAT'S NEXT?

Possible areas for future research include trialling the new tool in other countries – including leveraging the tool's cost advantages for low and middle-income countries – and testing the tool against other methods for creating EQ-5D-5L value sets.

The tool can also be adapted to create value sets for other health descriptive systems. The researchers are also keen to work with medical specialists who see value in applying the tool to personalised medicine.

QUESTIONS TO CONSIDER

1. What is a QALY?
2. Do you think that the five dimensions included in the EQ-5D-5L presented in Figure 1 adequately represent health-related quality of life (HRQoL)? Are there any dimensions missing, in your opinion?
3. With references to Figure 1, what combinations of the five dimensions (i.e. health states) would be worse than dead, in your opinion?

USEFUL WEBSITE

Conference presentations and a discussion paper about the tool: www.1000minds.com/about/news/eq5d-value-sets

REFERENCES

R. Brooks (1996), EuroQol: the current state of play, *Health policy*, 37, 53-72.

N. Devlin, P. Hansen, P. Kind & A. Williams (2003), Logical inconsistencies in survey respondents' health state valuations – A methodological challenge for estimating social tariffs, *Health Economics*, 12, 529-44.

G. Farin, J. Hoschek & M.-S Kim (2002), *Handbook of Computer Aided Geometric Design*, North Holland: Elsevier Science Ltd.

A Fitzgerald, C. de Coster, S. McMillan et al. (2011), Relative urgency for referral from primary care to rheumatologists: the Priority Referral Score, *Arthritis Care & Research*, 63, 231-39.

P. Hansen & F. Ombler (2008), A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives, *Journal of Multi-Criteria Decision Analysis*, 15, 87-107.

P. Hansen, A. Hendry, R. Naden, F. Ombler & R. Stewart (2012), A new process for creating points systems for prioritising patients for elective health services, *Clinical Governance: An International Journal*, 17, 200-209.

M. Herdman, C. Gudex, A. Lloyd et al. (2011), Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L), *Quality of Life Research*, 20, 1727-36.

N. Martelli, P. Hansen, H. van den Brink et al. (2016), Combining multi-criteria decision analysis and mini-health technology assessment: A funding decision-support tool for medical devices in a university hospital setting, *Journal of Biomedical Informatics*, 59, 201-208.

R. Mirnezami, J. Nicholson & A. Darzi (2012), Preparing for precision medicine, *New England Journal of Medicine*, 366, 489-91.

C. Shiboski, S. Shiboski, R. Seror et al. (2017), 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome: A consensus and data-driven methodology involving three international patient cohorts, *Arthritis & Rheumatology*, 69, 35-45.

T. Sullivan & P. Hansen (2017), Determining criteria and weights for prioritizing health technologies based on the preferences of the general population: A New Zealand pilot study, *Value in Health*, 20, 679-86.

E. Tacconelli, E. Carrara, A. Savoldi et al. (2018), Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis, *The Lancet Infectious Diseases*, 18, 318-27.

6 An article fully describing the tool and presenting the results from the survey is currently under review at an academic journal. The social value set for NZ is available on request from the authors.

7 This demonstration version of the tool is intended for a somewhat 'knowledgeable' audience (as you will see at its conclusion). The tool developed for patients has more user-friendly instructions, simplified results, and finishes with questions for collecting demographic information too.